DOCUMENT RESUME

ED 274 715 TM 860 594

AUTHOR Morris, John D.; Huberty, Carl J.
TITLE A Comparison of Three Methods of Classification Hit-Rate Estimation.
PUB DATE Apr 86
NOTE 13p.; Paper presented at the Annual Meeting of the American Educational Research Association (67th, San Francisco, CA, April 16-20, 1986).
PUB TYPE Speeches/Conference Papers (150) -- Reports - Research/Technical (143)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS *Classification; Comparative Analysis; *Correlation; *Estimation (Mathematics); Mathematical Formulas; Meta Analysis; Statistical Analysis; Statistical Studies
IDENTIFIERS *Cross Validation

ABSTRACT

Formulas for estimating cross-validated hit-rates, the number of correct classifications into an a priori grouping structure, were examined. The following mathematical formulas were compared: McLachlan's formula estimator, two Snappin and Knoke smoothed formula estimators, and the analytic leave-one-out estimator. The R method was included as a benchmark, but not considered as a viable candidate because of its optimistic bias. Twenty-one data sets were taken from real classification studies from journal articles, presented papers, and textbooks. The classification methods were accomplished as specified with each data set assuming equal population covariance matrices, prior probabilities of group membership, and costs of misclassification. Results showed that in most of the data sets, the estimation methods gave very similar results. Most methods overestimated the R method hit rate with some data sets, which was not expected. The McLachlan method, which does not perform as a multiple regression formula does, gave estimates larger than that of the R-method. The leave-one-out (LOO) method appeared to be the most desirable. References and tables are provided. (GDC)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED274715

A Comparison of Three Methods of Classification Hit-Rate Estimation

John D. Morris

Institute for Research and Development in Teacher Education

Florida Atlantic University

Carl J. Huberty

University of Georgia

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

J.D. Morris

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

X This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

Paper presented at the annual meeting of the American Educational Research Association, April 1986, San Francisco.

TM 860 594

A Comparison of Three Methods of Classification Hit-Rate Estimation

John D. Morris
Institute for Research and Development in Teacher Education
Florida Atlantic University

Carl J. Huberty
University of Georgia

As in the estimation of the cross-validated multiple correlation from sample data, the estimation of cross-validated "hit-rates" (number of correct classifications into an a priori grouping structure) in classification studies has been of interest in the general field of discriminant analysis for many years. However, unlike in the case of multiple correlation where formula estimates (Huberty & Mourad, 1980) are numerous, the theoretical derivation of parametric formula estimates for cross-validation hit-rates is much more complicated.

Also, as with estimators of cross-validated multiple correlation, one must decide what kind of cross-validation is of interest. For example, is estimating the correlation that would obtain upon application of the sample generated weights to population data of primary interest, or is the accuracy that would result with the application of unknown population weights of primary concern. An additional important concern in the case of multiple regression studies is whether the researcher is interested in relative or absolute accuracy (Morris, 1986). However, as classification is based on absolute accuracy this question does not apply in this study.

The accuracy criterion in this paper will be that obtained upon classification of a replicate sample. Analogous to the still not uncommon practice of using the sample generated multiple correlation as an index of accuracy in this context, most classification studies naively use the classification of the calibration sample (that sample from which classification functions were calculated) as an index of classification accuracy. Unfortunately, also analogous to multiple correlation, is the positive bias of this method (Frank, Massey, & Morrison, 1965; Lachenbruch & Mickey, 1968; Michaelis, 1973). In the context of classification this method has been labeled and will be referred to in this paper as the "R" method.

Only in the case of two-group classification has there been any success in deriving formula estimates for cross-validated hit-rate. Several formulas for the two-group case do exist and have been evaluated (Lachenbruch & Mickey, 1968; McLachlan, 1975: Page, 1985; Snappin & Knoke, 1985) for the two-group case. From these results, the estimator due to McLachlan (1975) along with two new "smoothed" estimators due to Snappin and Knoke (1985) seem to hold the most promise, yet have not been extensively empirically evaluated. Although the formula for the McLachlan estimator is quite complex, it is calculable with a hand-held calculator. The smoothed estimator due to Snappin and Knoke (1985) is simpler to represent than that due to McLachlan (1975), but actually requires more computation.

McLachlan (1975) derived the following expression for the estimated hit rate, P_g for group g, where g = 1 or 2:

$$P_g = 1 - F(-D/2) - f(-D/2)\{(p - 1)/(Dn_g) + D[4(4p - 1) - D^2]/(32m) + (p - 1)(p - 2)/(4Dn_g^2) + (p - 1)[-D^3 + 8D(2p + 1) + 16/D]/(64mn_g) + D[3D^6 - 4D^4(24p + 7) + 16D^2(48p^2 - 48p - 53) + 192(-8p + 15)]/(12288m^2)\},$$

where F is the standard normal distribution function, i.e., F(-D/2) is the area to the "left" of -D/2, f is the standard normal density function, D is the Mahalanobis distance, p is the number of predictor variables, n_g is the number of subjects in group g, and $m = n_1 + n_2 - 2$. While the formula looks formidable, with patience, it is calculable with a hand-held calculator. Moreover, as the last term in the multiplier for f(-D/2) is usually very small, one may choose to ignored it making the formula even more tractable. This formula estimate, using the full expansion, will be referred to in this paper as the "M" method.

Snappin and Knoke (1985) offered two formulas that were intended to "smooth" the results that one would obtain from the R method by estimating "partial" misclassifications for those subjects for which the classification decision is a "toss up." The intended result, based on the work of Glick (1978), was a lessening of the variance and bias of the R method. In order to define those algorithms, the R classification method must first be introduced.

The usual two group classification statistic for a subject with score vector $\mathbf{X}_i$, originally introduced by Wald (1944), is $W(\mathbf{X}_i) = \{\mathbf{X}_i - (\bar{\mathbf{X}}_1 + \bar{\mathbf{X}}_2)/2\}'\mathbf{S}^{-1}(\bar{\mathbf{X}}_1 - \bar{\mathbf{X}}_2)$, with $\bar{\mathbf{X}}_1$, $\bar{\mathbf{X}}_2$, and $\mathbf{S}$ being the sample means and pooled covariance matrix estimate. With equal prior group membership probabilities and costs of misclassification, the "R" method then consists of merely calculating this statistic for each subject in the calibration sample and classifying according to the following rule:

If $W(\mathbf{X}_i)$ > 0, classify subject i into group 1
≤ 0, classify subject i into group 2.

Such a classification statistic offers little confidence in classifying those subjects with a $W(\mathbf{X}_i)$ that is near 0; these subjects are descriptively called "fence riders" by Huberty (1984).

Snappin and Knoke derived a "smoothing constant", b, which for group one is, $b = \{[(p + 2)(n_1 - 1) + n_2 - 1]/[n_1(n_1 + n_2 - p - 3)]\}^{1/2}$. Clearly one can obtain the b for classifying group two by switching n_1, and n_2. The smoothed classification error rate estimators introduced by Snappin and Knoke based upon this smoothing constant were the "uniform" smoothing function and the "normal" smoothing function, and were both labeled the "NS" method. These methods will be referred to as "NSU", and "NSN" respectively in this paper.

The uniform smoothing function is identical to the R method except for the treatment of "fence riding" subjects for whom $W(\mathbf{X}_i)$ is near zero. "Near" was defined as an interval of bD about zero, so that only subjects with $W(\mathbf{X}_i)$ scores between -bD and +bD are classified differently than in the R method. Specifically, the uniform smoothing function, $g_1(\mathbf{X}_i)$ is,

$g_1(\mathbf{X}_i) = 0$ if $W(\mathbf{X}_i) > bD$
$= 1$ if $W(\mathbf{X}_i) < -bD$
$= \{bD - W(\mathbf{X}_i)\}/(2bD)$ otherwise.

One can see then that while the contribution to estimated error for a subject i with $-bD < W(\mathbf{X}_i) < bD$ would be either 0 or 1 for the R method, the NSU method would give a

contribution somewhere between 0 and 1 depending on $W(X_i)$, b, and D.

The normal smoothing function was given as $g_2(X_i) = F[-W(X_i)]/(bD)]$, where F has been defined previously. Thus the NSN method is not parallel to the R method in the mechanics of estimating the degree of misclassification. The contribution to estimated error depends on the position on the horizontal axis for the normal curve of $-W(X_1)/(bD)$.

As in the R method, both the NSU and NSN functions are to be aggregated over the entire group to yield an estimated proportion of misclassifications such that $P_g = 1 - \sum g(X_i)/n_g$. Therefore, because each subject's contribution to error is summed, these smoothed estimators require more computation than a formula estimate such as that of McLachlan.

One non-parametric analytic estimator for cross-validation hit-rate is the "leave-one-out" (LOO) estimator (Huberty, 1984; Huberty and Mourad, 1980; Lachenbruch and Mickey, 1968; Mosteller and Tukey, 1968). In this method a subject is classified by applying the classification function(s) based on all subjects except the one being classified. This process is repeated "round-robin" for each subject with the overall classification accuracy used to estimate the accuracy that would obtain on replicate samples.

This "resampling" plan is similar in thrust to both "jackknifing" (although not identical, see Stone, 1974), and "bootstrapping" (Efron, 1979, 1983). The technique is also used with a continuous criterion in multiple regression but is usually called PRESS (Allen, 1971; Allen and Cady, 1982, p. 254).

There are several advantages to the "leave-one-out" method. First is its generality; it is a simple and powerful idea that is applicable to almost any classification problem regardless of such concerns as the number of classification groups. It can even be applied to estimate hit-rates for biased weighting strategies such as ridge regression classification (Morris & Huberty, 1985). Note that regardless of the formula estimates' degree of accuracy, they are only appropriate with ordinary least squares weighting.

Also, as the technique makes no parametric assumptions about the population, the researcher has little concern for violations of assumptions on application to real data. Indeed, to the degree that a sample's distribution is representative of the population's, the technique models its assumptions to fit the data. The distribution assumed is therefore implicitly whatever the sample manifests.

Theoretically, the technique is biased as the classifications are accomplished on N - 1 rather than N subjects. But for even the smallest sample that would be of respectable size (as judged by other criteria) for use in a classification study, the degree of bias is logically, as well as empirically (Snappin & Knoke, 1985), trivial.

An immediately apparent disadvantage of the method is the implied computational labor. It would seem that N matrices would need to be inverted in order to accomplish the leave-one-out cross-validation estimate. It turns out, however, that Lachenbruch (1968) utilized a matrix identity due to Bartlett (1951) to introduce a computational shortcut that avoids the many matrix inversions.

Since the classification function for a two-group problem when population covariance matrices and a priori probabilities of group membership and costs of misclassification are assumed equal is proportional to the parallel regression equation predicting the dichotomous criterion, it should be no surprise that a similar shortcut was derived for the multiple regression PRESS algorithm (Allen, 1971). Utilizing the shortcut in the case of multiple regression has been shown to

yield a speed advantage of as much as an order of magnitude with some data (Morris, 1984). Therefore, while the technique is clearly more computation intensive than a simple estimation formula, as in McLachlan's, or even the Snappin and Knoke estimators, the difference is not as much as would accrue with the naive approach of inverting N matrices.

Moreover, computation costs are becoming much less important as fast microcomputers are utilized more in statistical analysis. Using a microcomputer, the cost for computation time is essentially zero, except for loss of the use of the machine for other purposes during computation. With currently available and emerging multitasking operating systems, loss of the microcomputer during extensive computations will also soon be of no concern. Tukey's (1985) comments relating to our need to make sure that the statistical techniques we invent anticipate the incredible resources of computational power that we will have in the near future seem especially relevant to this concern.

Another disadvantage of the leave-one-out strategy is that it has been shown in simulations that in populations in which the parametric assumptions of the formula estimates are correct, the leave-one-out estimator tends to be somewhat inferior to the formula estimators (Lachenbruch & Mickey, 1968; Snappin & Knoke, 1984; Toussaint, 1974). This difference can be largely attributed to the extra variance created by the leave-one-out method, as the bias is inconsequential. However, one must note that real data never fit parametric assumptions exactly, and that the parametric formula estimates are sensitive to certain departures from normality (Snappin & Knoke, 1985). How accurately these formula estimators perform with real data distributions does not appear to have been investigated.

Purpose

The purpose of this study was to compare the performance of the McLachlan (1975) formula estimator, the two Snappin & Knoke (1985) smoothed formula estimators, and the analytic leave-one-out estimator on several data sets taken from actual classification studies. Also included as a benchmark, rather than a viable candidate because of its optimistic bias, was the R method. The evidence suggests that a formula estimator is superior when parametric assumptions are met, however as with any real data, the data sets used in this paper do not exactly meet the assumptions. Yet a data set's distribution doesn't violate any assumptions accorded to the leave-one-out strategy; it should still be an almost unbiased estimator. Therefore, the degree to which the easier to calculate estimates mimic the leave-one-out estimator is the major question of concern.

Method

The data sets

Twenty-one data sets have been analyzed at this time. To bolster validity, all were taken from real classification studies. The sources are journal articles, paper presentations, and textbooks. Some of the data sets are "well known" and some are not. No real attempt was made to screen the data; we have used the data sets as they have become available. No pathological distributional problems are known in any of the data sets; it is hoped that the distributions are as one would find in typical classification studies in the literature. On the other hand, the results were not intended to and should not be interpreted as generalizable to all data sets. The intent was simply to investigate the performance of the methods on real data sets as

this question has not yet been addressed.

As these data sets have been "used" in the literature, it is assumed that any aberrant score vectors have been removed. In general it is expected that before applying the leave-one-out strategy (or any other) that one would wish to consider the removal of "outliers" that manifest appreciable leverage. See the excellent review by Hocking (1983), and associated comments for information on methods for detecting such score vectors in the case of multiple regression. Because of the parallel between methods, the results should also apply in the case of two-group classification.

Classification

The classification methods were accomplished as specified with each data set assuming equal population covariance matrices, prior probabilities of group membership, and costs of misclassification.

Results

For each of the data sets, Table 1 gives a short description (more information is available on request), the number of subjects classified, the relevant Mahalanobis D illustrating the degree of group separation, and the estimated proportion of hits by each of the five methods for group one. Table 2 gives the same information for group two.

Tables 1 and 2 will be discussed together as the ordering of groups is arbitrary. For the two classification problems with very large group separation (approximately D > 10) no distinction was available among methods as they all estimated no classification error.

In many other data sets the estimation methods gave very similar results. This is even true for the previously documented optimistically biased R method. There were notable exceptions, however, as in data sets 12 and 21 (Table 2) in which the NSU and NSN methods estimated far greater accuracy than either the LOO or M methods. Ultimately, however, the decision of how much difference is important must be decided by the individual researcher.

It is interesting to note that all contending methods (not including the R method as it was included only as a benchmark) except LOO overestimated the R method hit-rate with some data sets. In many cases, this overestimation was small, less than one percent. This may seem inconsequential, and it may be, but one should note that as the R method provides an optimistically biased estimate, one would expect that these other estimators would yield estimates that are smaller than those from the R method.

Yielding an estimate that is larger than the known optimistically biased sample estimate is quite different than one might expect from a multiple regression perspective. Parallel formula estimators of the cross-validated multiple correlation, or functions of it, such as those due to Nicholson (1948), Lord (1950), Darlington (1968), Rozeboom (1978), or Browne (1975) are functions of the sample multiple correlation, and are therefore guaranteed to yield smaller values than the sample value. The one formula estimator for hit-rate examined in this paper due to McLachlan is not quite parallel to multiple regression formula estimators. While it is an estimator of cross-validated hit-rate, it is not a function of the calibration sample generated hit-rate, rather, it is a function of the distance between groups as well as other variables. That is, it does not simply estimate a parameter from a function of the corresponding statistic as do the multiple regression formula

estimators. It is therefore seen to give estimates of hit-rate that are in some cases larger (up to 4%) than that of the R method (Table 1, data sets 3, 4, 16, & 18).

Although not a formula estimator, the NSU method of Snappin and Knoke is in some sense a function of the classification accuracy of the calibration sample, with the intent of reducing the bias and variance of that accuracy estimate. Nonetheless, it also gave hit-rate estimates for many samples (Table 1, data sets 6, 7, 8, 11, 12, 13, 16, 18, & 19; Table 2, data sets 4, 5, 6, 15, 16, 17, 19, & 20) that were larger than those obtained through the positively biased R method. These estimates ranged up to about seven percent larger than those obtained with the R method.

In the NSU method, all subjects except those for which $-bD < W(\mathbf{X}_i) < bD$, yield the same contribution to estimated hit-rate as in the R method. As mentioned previously, the remaining subjects are assigned a "contribution" to estimated error of somewhere between 0 and 1, depending on their $W(\mathbf{X}_i)$ score, b, and D. For these same cases, the R method gives an estimated error equal to the number of cases for which $W(\mathbf{X}_i) < 0$. Considering only the relevant subjects for which $-bD < W(\mathbf{X}_i) < bD$, whether the R or NSU method gives a larger estimated error is determined by the degree of balance between those subjects with $W(\mathbf{X}_i) > 0$ for which the NSU contribution is larger than that for the R method, and those subjects for which $W(\mathbf{X}_i) < 0$ for which the NSU contribution is less. This appears to be specific to the data set, but as can be seen, the NSU method gave estimated errors larger than the R method in nearly half the data sets.

Using the NSN method, the situation was only slightly improved with the number of data sets in which it estimated a larger hit-rate than the R method being nine.

Only in the case of the LOO method does the researcher seem to be assured of obtaining estimated error that is less than or equal to that estimated by the positively biased R method. Even in the case of the LOO method, however, the R method hit-rate was equaled in some data sets.

If we consider the LOO method to be the most desirable (although more computationally difficult) estimate because of the desirable characteristics mentioned earlier, and because of the overestimation of the other methods, a question of possible interest is how accurately these other estimators "track" the LOO method. Using the LOO method as the criterion, the respective correlations for each method as a predictor were .91 (R), .97 (M), .93 (NSU), and .94 (NSN).

As they are so close, these correlations really offer little discrimination among methods. One conclusion which one might come to is that it really doesn't matter which method is used. However, the limited generality of these data sets certainly needs to be considered in such a decision. Moreover, that there is in general fairly high agreement does not protect the researcher from an estimate that is even more positively biased than that due to the R method. As well, even though there is relatively high agreement among methods over all data sets, there are data sets in which none of the contenders are very close to LOO (Table 1, data sets 8, 12, 14, 17, 19, 20, & 21; Table 2, data sets 12 & 18).

An important question that researchers often ask about classification results is whether they are more accurate than would be expected by chance. (See Huberty [1984] for a thorough presentation of the issues and procedures involved.) Therefore, researchers may ask this question of the hit-rates derived from the estimator chosen. Another way of comparing the performance of these estimators then is to consider whether the researcher would be led to different decisions concerning the

"significance" of the accuracy of an available classification function based on the different hit-rate estimators. Again, contrasting performance with the LOO strategy, and assuming the researcher chose either a .05 or .01 significance level, the significance of the individual group estimated hit-rate for each of the methods was often different from that of LOO for all methods. In the case of the M method, there were occasions in which significance would be obtained by LOO, but not by M, and vice versa. As the NSU and NSN tended to overestimate LOO, the disagreement was only in one direction; NSU and NSN were significant when LOO was not. As would be expected, the R method was often significant when the LOO method was not.

All computations were done using an MS DOS microcomputer. While a lot of computation takes place, there were no runs of more than a few minutes. For this reason, it could be suggested that one could simply obtain all of these estimators for a particular data set. Toward this purpose, this program is available as a .COM file for anyone wishing it. The program is available for 8088 microcomputers, or for microcomputers that can also utilize an 8087 APU; please specify which is desired. If the program is of interest, send a blank DSDD diskette to:

John D. Morris
IRDTE
College of Education
Florida Atlantic University
Boca Raton, Florida 33431

References

Allen, D. M. (1971). The prediction sum of squares as a criterion for selecting predictor variables (Tech. Rep. No. 23). University of Kentucky, Department of Statistics.

Allen, D. M., & Cady, F. B. (1982). Analyzing experimental data by regression. Belmont, CA: Wadsworth.

Bartlett, M. S. (1951). An inverse matrix adjustment arising in discriminant analysis. Annals of Mathematical Statistics, 22, 107.

Browne, M. W. (1975). Predictive validity of a linear regression equation. British Journal of Mathematical and Statistical Psychology, 28, 79-87.

Darlington, R. B. (1968). Multiple regression in psychological research and practice. Psychological Bulletin, 69, 161-182.

Efron, B. (1979). Bootstrap methods: Another look at the jackknife. Annals of Statistics, 7, 1-26.

Efron, B. (1983). Estimating the error of a prediction rule: Improvement on cross-validation. Journal of the American Statistical Association, 78, 316-331.

Frank, R. E., Massey, W. F., & Morrison, D. F. (1965). Bias in multiple discriminant analysis. Journal of Marketing Research, 2,250-258.

Glick, N. (1978). Additive estimators for probabilities of correct classification. Pattern recognition, 10, 211-222.

Hocking, R. R. (1983). Developments in linear regression methodology: 1959-1982. Technometrics, 25, 219-230.

Huberty, C. J. (1984). Issues in the use and interpretation of discriminant analysis. Psychological Bulletin, 95, 156-171.

Huberty, C. J., & Mourad, S. A. (1980). Estimation in multiple correlation/ prediction. Educational and Psychological Measurement, 40, 101-112.

Lachenbruch, P. A., & Mickey, M. R. (1968). Estimation of error rates in discriminant analysis. Technometrics, 10, 1-11.

Lord, F. M. (1950). Efficiency of prediction when a regression equation from one sample is used in a new sample. (Research Bulletin No. 50-40) Princeton, N.J.: Educational Testing Service.

Michaelis, J. (1973). Simulation experiments with multiple group linear and quadratic discriminant analysis. In T. Cacoullos (Ed.), Discriminant analysis and applications. New York: Academic Press.

McLachlan, G. J. (1975). Confidence intervals for the conditional probabilities of misallocation in discriminant analysis. Biometrics, 31, 161-167.

Morris, J. D. (1984). Cross-validation with Gollob's estimator: A computational simplification. Educational and Psychological Measurement, 41, 151-154.

Morris, J. D., & Huberty, C. J. (1985, April). On selecting a classification weighting algorithm. Paper presented at the meeting of the American Educational Research Association, Chicago.

Morris, J. D. (1986, in press). Selecting a predictor weighting method by PRESS. Educational and Psychological Measurement.

Mosteller, F., and Tukey, J. W. (1968). Data analysis, including statistics. In G. Lindzey & E. Aronson (Eds.), Handbook of Social Psychology: Vol. 2. Reading Mass.: Addison-Wesley.

Nicholson, G. E. (1948). The application of a regression equation to a new sample. Unpublished doctoral dissertation, University of North Carolina.

Page, J. T. (1985). Error-rate estimation in discriminant analysis. Technometrics, 27, 189-198.

Rozeboom, W. W. (1978). Estimation of cross-validated multiple correlation: A clarification. Psychological Bulletin, 85, 1348-1351.

Snappin, S. M., & Knoke, J. D. (1984). Classification error rate estimators in discriminant analysis. Technometrics, 26, 371-378.

Snappin, S. M., & Knoke, J. D. (1985). An evaluation of smoothed classification error-rate estimators. Technometrics, 27, 199-206.

Stone, M. (1974). Cross-validatory choice and assessment of statistical predictions. Journal of the Royal Statistical Society, Series B, 36, 111-147.

Toussaint, G. T. (1974). Bibliography on estimation of classification. IEEE Transactions on Information Theory, 20 472-479.

Tukey, J. W. (1985). Comment. The American Statistician, 39, 12-14.

Wald, A. (1944). On a statistical problem arising in the classification of an individual into one of two groups. Annals of Mathematical Statistics, 15, 145-162.

Table 1

Proportion of Correct Classifications Estimated by each Method for First of Two Groups*

Numerical Designator and Data Set Description	n	D	Method R	As a % of R: M	NSU	NSN	LOO
1 Fisher Groups 1 & 3	50	13.97	1.0000	100.0	100.0	100.0	100.0
2 Fisher Groups 1 & 2	50	10.16	1.0000	100.0	100.0	100.0	100.0
3 Bisbey Groups 1 & 3	35	5.12	.9714	101.2	99.9	99.7	97.1
4 Fisher Groups 2 & 3	50	3.77	.9600	100.6	100.0	99.5	100.0
5 Bisbey Groups 1 & 2	35	2.89	.9143	96.2	98.8	98.0	96.9
6 Rulon Groups 1 & 3	85	2.86	.9529	95.8	100.3	99.8	100.0
7 Bisbey Groups 2 & 3	81	2.41	.8889	95.0	100.1	98.2	93.1
8 Talent Groups 3 & 5	33	1.97	.8182	91.8	104.2	103.1	96.3
9 Rulon Groups 2 & 3	93	1.87	.8602	95.1	98.7	97.9	100.0
10 Rulon Groups 1 & 2	85	1.73	.8471	94.3	98.9	97.7	100.0
11 Talent Groups 1 & 5	83	1.72	.7952	95.4	101.0	100.6	92.4
12 Talent Groups 1 & 3	83	.89	.6867	85.3	100.5	99.1	93.0
13 Block Groups 3 & 4	38	.85	.7368	84.3	100.6	98.5	100.0
14 Block Groups 1 & 2	40	.84	.6750	92.1	94.8	93.2	85.2
15 Block Groups 1 & 4	40	.81	.6250	98.5	94.8	93.7	92.0
16 Block Groups 1 & 3	40	.74	.5750	104.3	100.4	100.8	100.0
17 Warncke Groups 1 & 3	65	.68	.7231	74.9	96.9	93.8	85.1
18 Block Groups 2 & 3	37	.64	.5676	100.3	103.6	102.8	100.0
19 Block Groups 2 & 4	37	.52	.6486	82.8	101.5	100.9	95.8
20 Warncke Groups 1 & 2	65	.48	.6154	76.3	98.1	98.1	82.5
21 Warncke Groups 2 & 3	47	.44	.6596	60.8	94.7	92.4	67.7

*Note that ordering of groups is arbitrary.

Table 2

Proportion of Correct Classifications Estimated by each Method for Second of Two Groups*

Numerical Designator and Data Set Description	n	D	Method R	Method As a % of R: M	NSU	NSN	LOO
1 Fisher Groups 1 & 3	50	13.97	1.0000	100.0	100.0	100.0	100.0
2 Fisher Groups 1 & 2	50	10.16	1.0000	100.0	100.0	100.0	100.0
3 Bisbey Groups 1 & 3	37	5.12	1.0000	98.4	100.0	99.7	100.0
4 Fisher Groups 2 & 3	50	3.77	.9800	98.5	100.4	99.6	100.0
5 Bisbey Groups 1 & 2	81	2.89	.9136	97.7	100.2	99.6	97.3
6 Rulon Groups 1 & 3	67	2.86	.9254	99.3	100.3	99.4	100.0
7 Bisbey Groups 2 & 3	37	2.41	.8919	92.7	99.1	98.2	97.0
8 Talent Groups 3 & 5	94	1.97	.8085	97.8	99.3	99.2	94.7
9 Rulon Groups 2 & 3	66	1.87	.8182	99.9	98.6	97.6	98.1
10 Rulon Groups 1 & 2	93	1.73	.8172	97.8	100.0	99.2	100.0
11 Talent Groups 1 & 5	94	1.72	.7872	96.8	99.4	99.1	95.9
12 Talent Groups 1 & 3	33	.89	.7879	60.3	93.4	89.1	53.8
13 Block Groups 3 & 4	38	.85	.6842	90.8	98.8	95.1	92.3
14 Block Groups 1 & 2	37	.84	.8108	76.3	97.6	93.3	93.3
15 Block Groups 1 & 4	38	.81	.6579	93.3	103.0	101.1	96.0
16 Block Groups 1 & 3	38	.74	.7368	81.1	100.8	98.5	96.4
17 Warncke Groups 1 & 3	40	.68	.5750	85.1	106.8	106.1	87.0
18 Block Groups 2 & 3	38	.64	.6316	90.3	96.0	95.1	83.3
19 Block Groups 2 & 4	38	.52	.5526	97.4	104.5	106.3	100.0
20 Warncke Groups 1 & 2	47	.48	.6170	68.5	103.5	100.8	72.4
21 Warncke Groups 2 & 3	40	.44	.6750	54.6	98.6	95.2	51.9

*Note that ordering of groups is arbitrary.

END

U.S. DEPT. OF EDUCATION

OFFICE OF EDUCATIONAL
RESEARCH AND
IMPROVEMENT (OERI)

ERIC

DATE FILMED

FEB_24_1987